Комбинаторные оценки вероятности переобучения пороговых конъюнкций для логических алгоритмов классификации

Ивахненко А. А.

1 октября 2010 г.

Аннотация

Получены и исследованы комбинаторные оценки вероятности переобучения для логических правил, имеющих вид пороговых конъюнкций над заданным подмножеством вещественных признаков. Введено понятие фиксированных объектов и предложены алгоритмы их эффективного вычисления. С их помощью построены верхние оценки вероятности переобучения, учитывающие эффекты расслоения и связности в семействе пороговых конъюнкций. Эти оценки предлагается использовать в качестве критерия информативности при поиске конъюнктивных закономерностей в логических алгоритмах классификации.

1 Задача индукции логических правил

Пусть задана выборка объектов $\mathbb{X} = (x_i)_{i=1}^L$, описанных n действительными признаками, $x_i = (x_i^1, \ldots, x_i^n)$, и каждому объекту x_i соответствует ответ y_i из заданного конечного множества Y.

Алгоритмы классификации $a\colon \mathbb{R}^n \to Y$, основанные на взвешенном голосовании логических правил (rules), имеют следующий вид:

$$a(x) = \arg\max_{y\in Y} \sum_{r\in R_y} w_r r(x),$$

где w_r — вес правила r, обычно неотрицательный, R_y — множество правил класса y. В общем случае *правило* — это функция вида $r\colon \mathbb{R}^n \to \{0,1\}$ из некоторого фиксированного параметрического семейства R. В данной работе рассматривается один из наиболее распространённых типов правил — семейство конъюнкций пороговых предикатов:

$$r(x) \equiv r(x; c^1, \ldots, c^n) = \prod_{j\in\omega} \big[x^j \lessgtr_j c^j\big], \quad (1)$$

где $x = (x^1, \ldots, x^n) \in \mathbb{R}^n$, $\omega \subseteq \{1, \ldots, n\}$ — подмножество признаков, $\lessgtr_j$ — одна из операций сравнения $\{\leqslant, \geqslant\}$, c^j — порог по j-му признаку.

Говорят, что правило r *выделяет* объект x, если $r(x) = 1$. Предполагается, что правила класса y должны выделять как можно больше объектов класса y и как можно меньше объектов всех остальных классов. Поэтому для поиска (*индукции*) правил класса y по обучающей выборке $X \subset \mathbb{X}$ решается задача двухкритериальной оптимизации:

$$P(r, X) = \sum_{x_i\in X} r(x_i)\,[y_i = y] \to \max_r;$$
$$N(r, X) = \sum_{x_i\in X} r(x_i)\,[y_i \neq y] \to \min_r;$$

На практике для этого оптимизируют некоторый критерий информативности, который является функцией от пары исходных критериев (P, N). В частности, это может быть энтропийный критерий, индекс Джини, точный тест Фишера, тест χ^2, тест ω^2 и другие [4], однако ни один из них не является безусловно предпочтительным. Большинство критериев оценивают степень неслучайности разбиения обучающей выборки X на два подмножества (положительные примеры $x\colon r(x) = 1$ и отрицательные $x\colon r(x) = 0$) относительно исходного разбиения выборки X на классы.

Недостаток стандартных критериев информативности в том, что они не учитывают переобучение. При оптимизации $P(r, X)$ и $N(r, X)$

по обучающей выборке X соответствующие величины $P' = P(r, \bar{X})$ и $N' = N(r, \bar{X})$ уже не будут оптимальны на контрольной выборке $\bar{X} = \mathbb{X}\backslash X$. Оценки [2, 5], основанные на теории Вапника-Червоненкиса [1], позволяют связать вероятность переобучения со сложностью семейства правил R. Они зависят от ранга конъюнкции $|\omega|$ и числа допустимых значений признаков x^j по каждой размерности $j \in \omega$, но не зависят от конкретной выборки данных $\mathbb{X}$. Согласно экспериментам [5] эти оценки сильно завышены, что делает их непригодными для количественного предсказания значений (P', N').

Точные комбинаторные оценки вероятности переобучения [6, 7] позволяют учитывать свойства расслоения и связности — более тонкие характеристики семейства, зависящие от данных $\mathbb{X}$. До сих пор точные оценки удавалось получать лишь для некоторых модельных семейств алгоритмов: монотонных и унимодальных цепочек и многомерных сеток, интервалов и шаров булева куба, и т. п.

В данной работе предлагается оценка для семейства правил (1), используемого при решении практических задач [3]. Цель работы — получить критерий информативности набора признаков ω, который учитывал бы величину переобучения, возникающего при оптимизации порогов c^j, $j \in \omega$.

2 Вероятность переобучения правил

Произвольное правило $r \in R$ класса $y \in Y$ индуцирует на выборке $\mathbb{X}$ бинарный вектор ошибок $\vec{r} = (r_i)_{i=1}^L$, где $r_i = \big[r(x_i) \neq [y_i{=}y]\big]$.

Определим *число ошибок* правила r на выборке $X \subseteq \mathbb{X}$

$$m(r, X) = \sum_{x_i \in X} r_i.$$

и *частоту ошибок* правила r на выборке X

$$\nu(r, X) = \frac{1}{|X|} m(r, X).$$

Методом обучения называется отображение вида $\mu\colon 2^{\mathbb{X}} \to R$, которое произвольной обучающей выборке $X \subseteq \mathbb{X}$ ставит в соответствие некоторое правило $r = \mu X$ из R. Метод μ называется методом *минимизации эмпирического риска* (МЭР), если

$$\mu X \in R(X) = \operatorname{Arg}\min_{r \in R} m(r, X);$$

и *пессимистичным* методом МЭР, если

$$\mu X = \arg\max_{r \in R(X)} m(r, \mathbb{X});$$

Пессимистичный метод МЭР не реализуем на практике, т. к. контрольная выборка скрыта в момент обучения. Тем не менее, он представляет значительный теоретический интерес, так как точные оценки вероятности переобучения для пессимистичного метода МЭР являются достижимыми верхними оценками для произвольного метода МЭР.

Разобъём выборку $\mathbb{X}$ всеми C_L^ℓ способами на две непересекающиеся подвыборки: обучающую X длины ℓ и контрольную $\bar{X}$ длины $k = L - \ell$, неизвестную в момент обучения. Следуя слабой вероятностной аксиоматике [5], сделаем единственное вероятностное предположение, что все C_L^ℓ разбиений равновероятны. Данное предположение фактически эквивалентно стандартному предположению о независимости наблюдений в двух подвыборках. Определим для любого $\varepsilon > 0$ *вероятность переобучения* метода μ как

$$Q_\varepsilon(\mu, \mathbb{X}) = \mathsf{P}\big[\nu(\mu X, \bar{X}) - \nu(\mu X, X) > \varepsilon\big], \quad (2)$$

где знак вероятности можно понимать как среднее по всем разбиениям: $\mathsf{P} \equiv \frac{1}{C_L^\ell} \sum_{X \sqcup \bar{X}}$.

Возьмём в качестве R семейство правил (1) с полным набором признаков $\omega = \{1, \ldots, n\}$ и операцией сравнения $\leqslant$ по всем признакам. Без ограничения общности будем полагать, что все правила относятся к фиксированному классу $y \in Y$. Нашей основной задачей будет получение оценок вероятности переобучения для данного семейства.

3 Структура классов эквивалентности правил

Два правила эквивалентны, $r \sim r'$, если их векторы ошибок совпадают, $\vec{r} = \vec{r}'$.

Допустим, что значения x_i^j каждого признака $j \in \omega$ попарно различны на объектах выборки $\mathbb{X}$. Это предположение будет, в частности, выполнено с вероятностью 1, если объекты x_i выбирались из непрерывного распределения на $\mathbb{R}^n$.

Пусть $x_j^{(i)}$ — i-й элемент в вариационном ряду значений j-го признака $x_j^{(1)} < \cdots < x_j^{(L)}$. Заменим в исходной матрице данных $\|x_i^j\|_{L\times n}$ каждое значение $x_j^{(i)}$ его рангом i в вариационном ряду. Очевидно, на этой матрице данных семейство R индуцирует то же множество векторов ошибок, что и на исходной. Поэтому далее будем полагать, что все признаки принимают целые значения $1,\ldots,L$, и никакие два объекта не имеют равных значений одного и того же признака. Значения порогов c^j в правилах $r(x;c^1,\ldots,c^n)$ вида (1) также имеет смысл выбирать только из целых значений $0,\ldots,L$

Будем говорить, что правила *связаны*, если их векторы ошибок различаются только на одном объекте. Заметим, что число классов эквивалентности и граф связей зависят от значений признаков $\|x_i^j\|$, но не зависят от классификаций y_i.

Пусть u и v — произвольные объекты из $\mathbb{X}$. Будем говорить, что объект u *доминируется* объектом v по координате j, если $u^j < v^j$. Будем говорить, что объект u *доминируется* множеством объектов $S \subseteq \mathbb{X}$ и писать $u \prec S$, если для каждого $j \in \omega$ существует объект $s \in S$, такой, что $u^j < s^j$. Если множество S состоит из одного элемента s и $u \prec S$, то будем записывать $u \prec s$.

Опр. 3.1 *Подмножество $S \subseteq \mathbb{X}$ называется* недоминирующимся, *если любой объект $s \in S$ не доминируется подмножеством $S\backslash s$.*

Обозначим множество всех недоминирующихся подмножеств мощности q через M_q,

$$M_q = \big\{S \subseteq \mathbb{X}\colon |S| = q,\ \forall s \in S\ \ s \not\prec S\backslash s\big\}.$$

Введем искусственное недоминирующееся подмножество S_0, состоящее из одного объекта $x_0 = (0,\ldots,0)$. Обозначим $M_0 = \{S_0\}$. Правило $r(x;0,\ldots,0)$ будем обозначать r_0.

Мощность произвольного недоминирующегося подмножества $|S|$ не превышает n. Если подмножество S недоминирующееся, то любое его подмножество $S' \subset S$ также недоминирующееся. Очевидно, $|M_q| \leqslant C_L^q$. Для построения всех $S \in M_q$ достаточно добавить к каждому подмножеству $S' \in M_{q-1}$ один объект, еще не входящий в него; если полученное подмножество недоминирующееся, то оно войдет в M_q.

Лемма 3.1 *Для любого объекта x из недоминирующегося подмножества S найдется хотя бы один признак $j \in \omega$, по которому на данном x достигается $\max\limits_{s\in S} s^j$, причём*

$$\bigcup_{j=1}^{n} \operatorname{Arg}\max_{s\in S}(s^j) = S.$$

Доказательство.

Допустим, что это не так. Тогда существует $x \in S$, такой что для любого $j \in \omega$ существует $s \in S\backslash x$, для которого $x^j < s^j$. Следовательно $x \prec S\backslash x$, значит, S не является недоминирующимся подмножеством. ■

Поставим в соответствие подмножеству S правило

$$r(x,S) = r(x;\max_{x\in S} x^1,\ldots,\max_{x\in S} x^n).$$

Очевидно, разным S ставятся в соответствие разные $r(x,S)$, т. к. значения каждого признака $x_i^j,\ \ i = 1,\ldots,L$ попарно различны. Следовательно, зная j и x_i^j, можно однозначно указать объект x_i. Следовательно, набор параметров $c^j = \max\limits_{x\in S} x^j,\ \ j = 1,\ldots,n$ задаёт подмножество S однозначно, и различным S не могут соответствовать одинаковые $r(x,S)$.

На рис. 1 показан пример задачи с $n = 2$ признаками, $L = 10$ объектами и семейство R правил вида

$$r(x;c^1,c^2) = [x^1 \leqslant c^1]\,[x^2 \leqslant c^2].$$

Каждое правило из R задается парой порогов (c^1,c^2), поэтому правилам соответствуют узлы прямоугольной сетки $H = \{0,\ldots,L\}^2$. Отрезками соединены правила, лежащие в одном классе эквивалентности. Рядом с каждым классом эквивалентности подписано число ошибок на полной выборке $m(r,\mathbb{X})$. Все одноэлементные подмножества являются недоминирующимися. Среди двухэлементных подмножеств недоминирующимися являются

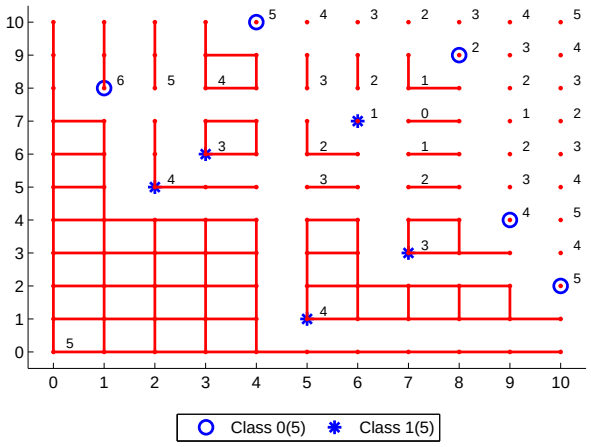

Рис. 1: Двумерная выборка длины $L = 10$, по 5 объектов в каждом классе (отмечены крупными точками).

только пары несравнимых объектов, например, пара объектов $\{(2,5),(5,1)\}$. Недоминирующихся подмножеств мощности больше 2 нет, т. к. размерность задачи $n = 2$.

Лемма 3.2 *Пусть $E \subset R$ — класс эквивалентности правил, $c^j(r)$ — j-й параметр правила r. Тогда классу E принадлежит также и правило*

$$r_E(x) = r\big(x; \min_{r' \in E} c^1(r'), \ldots, \min_{r' \in E} c^n(r')\big).$$

Доказательство.

В силу эквивалентности и бинарности всех правил $r' \in E$ правило

$$r(x) = \prod_{r' \in E} r'\big(x; c^1(r'), \ldots, c^n(r')\big)$$

принимает на всех объектах $x \in \mathbb{X}$ те же значения, $r(x) = r'(x)$, что и любое правило r' из E. Кроме того, r представимо в виде (1):

$$r(x) = \prod_{r' \in E} \prod_{j \in \omega} [x^j \leqslant c^j(r')] = \prod_{j \in \omega} [x^j \leqslant \min_{r' \in E} c^j(r')] = r_E(x).$$

Таким образом, правило $r_E(x)$ также принадлежит E. Что и требовалось доказать. ■

Будем называть правило $r_E(x)$ *стандартным представителем* класса эквивалентности E. На рис. 1 стандартные представители соответствуют левым нижним точкам каждого класса эквивалентности: $(0,0)$, $(1,8)$, $(2,5)$, $(5,1)$, и т. д.

Теорема 3.3 *Существует взаимно однозначное соответствие между множеством всех классов эквивалентности и множеством всех недоминирующихся подмножеств.*

Доказательство.

Построим по заданному стандартному представителю $r_E(x)$ недоминирующееся подмножество S. Если $r_E(x) = r_0$, то $S = S_0$. Иначе пусть $r_E(x) = r(x; c^1, \ldots, c^n)$. Рассмотрим множество объектов $U = \{x \colon r_E(x) = 1\}$. Оно не пусто, так как в противном случае класс эквивалентности E содержал бы r_0. Выберем из множества U недоминирующееся подмножество

$$S = \bigcup_{j=1}^{n} \operatorname{Arg}\max_{x \in U}(x^j).$$

Пусть $r(x,S) = r(x; d^1, \ldots, d^n)$. Докажем, что $r_E(x) = r(x,S)$ от противного. Допустим, что некоторые пороги в правилах $r_E(x)$ и $r(x,S)$ не совпадают; скажем, для определенности, $d^1 < c^1$. Возьмем правило $r(x; d^1, c^2, \ldots, c^n)$. Оно принадлежит E, так как $d^1 = \max\limits_{x \in U} x^1$, и при фиксированных параметрах $c^2, \ldots, c^n$ нет объекта, на котором, при увеличении первого порога от d^1 до c^1, изменялся бы вектор ошибок. Следовательно, $r_E(x)$ не может быть стандартным представителем класса эквивалентности E. Получили противоречие.

Теперь докажем обратное: любому недоминирующемуся подмножеству объектов S соответствует один и только один класс эквивалентности E. Поставим в соответствие недоминирующемуся подмножеству S класс эквивалентности E, такой, что $r(x,S) \in E$. Покажем от противного, что не существует второго такого S', $S' \neq S$, что $r(x,S') \in E$. Пусть это не так: $r(x,S) = r(x; c^1, \ldots, c^n) \in E$, $r(x,S') = r(x; d^1, \ldots, d^n) \in E$ и, для определенности, $d^1 < c^1$. Но тогда существует объект $x_* = \arg\max\limits_{x \in S}(x^1)$, такой, что $x^1_* = c^1$. При этом $r(x_*, S) = 1$ и $r(x_*, S') = 0$, следовательно эти два правила не могут принадлежать одному классу эквивалентности. ■

Следствие. Для каждого класса эквивалентности E существует единственное недоминирующееся подмножество S, такое, что $r_E(x) = r(x, S)$.

Следствие. Число классов эквивалентности равно $\sum_{q=0}^{n} |M_q|$.

4 Оценка вероятности переобучения

В методе порождающих и запрещающих множеств [7] вероятность переобучения $Q_\varepsilon(\mu, \mathbb{X})$ может быть определена точно, если для каждого правила $r \in R$ указана совокупность подмножеств $\{X_{rv}, X'_{rv} \subset \mathbb{X} \colon v \in V_r\}$ и коэффициенты c_{rv}, такие, что для любой подвыборки $X \subset \mathbb{X}$ длины ℓ

$$[\mu X = r] = \sum_{v \in V_r} c_{rv} [X_{rv} \subseteq X][X'_{rv} \subseteq \bar{X}]. \quad (3)$$

Подмножества X_{rv} называются *порождающими*, X'_{rv} — *запрещающими*. Указать такие подмножества всегда возможно, но не единственным способом. Чем меньше мощности множеств V_r, X_{rv}, X'_{rv}, тем эффективнее будет вычисляться оценка Q_ε.

Опр. 4.1 *Множества объектов*

$$X_r = \bigcap_{v \in V_r} X_{rv}, \quad X'_r = \bigcap_{v \in V_r} X'_{rv}$$

называются фиксированными *для правила* r.

Фиксированные объекты обязаны присутствовать во всех порождающих и запрещающих множествах для того, чтобы метод обучения μ выбрал правило r по обучающей выборке X. Перепишем гипотезу (3) в виде верхней оценки:

$$[\mu X = r] \leqslant [X_r \subseteq X][X'_r \subseteq \bar{X}]. \quad (4)$$

Теорема 4.1 *Если справедливо* (4), *то*

$$Q_\varepsilon(\mu, \mathbb{X}) \leqslant \sum_{r \in R} P_r H_{L_r}^{\ell_r, m_r}(s_r(\varepsilon)), \quad (5)$$

где $m_r = m(r, \mathbb{X} \backslash X_r \backslash X'_r)$ — *число ошибок правила* r *на нефиксированных объектах;*

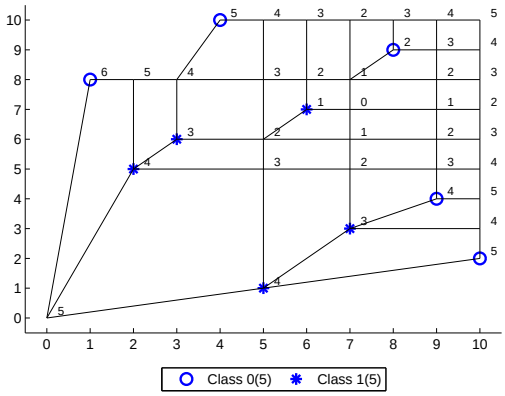

Рис. 2: Граф связей правил, показанных на рис. 1.

$L_r = L - |X_r \cup X'_r|$ *и* $\ell_r = \ell - |X_r|$ — *число нефиксированных объектов, соответственно, в полной выборке и в обучающей;*

$P_r = C_{L_r}^{\ell_r} / C_L^\ell$ — *верхняя оценка вероятности* $\mathsf{P}[\mu X = r]$ *получить правило* r *в результате обучения;*

$s_r(\varepsilon) = \frac{\ell}{L}(m(r, \mathbb{X}) - \varepsilon k) - m(r, X_r)$ — *максимальное число ошибок на нефиксированных обучающих объектах, при котором имеет место переобучение;*

$H_L^{l,m}(s) = \sum_{t=0}^{s} \frac{C_m^s C_{L-m}^{\ell-s}}{C_L^\ell}$ — *функция гипергеометрического распределения [7].*

Рассмотрим граф связей между правилами, рис. 2. Каждая вершина этого графа — это стандартный представитель класса эквивалентности, а связь между вершинами означает, что векторы ошибок правил различаются на одном объекте. Заметим, что топология графа зависит только от значений признаков объектов, но не зависит от классификации объектов.

Напомним, что рассматриваются только правила, относящиеся к фиксированному классу y. Будем говорить, что множество объектов $D(q, r) \subseteq \mathbb{X}$ *ухудшает* правило q по сравнению с правилом r, если:

1) $r(x_i) = q(x_i)$ для всех $x_i \in \mathbb{X} \backslash D(q, r)$;

2) $q(x_i) = [y_i = y]$, $r(x_i) = [y_i \neq y]$ для всех $x_i \in D(q, r)$.

Правила r и q различаются только на множестве $D(q, r)$, при этом правило r ошибается на всех объектах этого множества, а правило q — ни на одном. Если не существует такого

множества объектов для пары правил q и r, то будем полагать $D(q,r) = \varnothing$.

Теорема 4.2 *Пусть множество объектов $D(q,r)$ ухудшает правило r по сравнению с правилом q. Тогда, чтобы пессимистичный метод МЭР выбрал правило r, объекты множества $D(q,r)$ должны находиться в фиксированном множестве X'_r.*

Доказательство.

Докажем от противного. Пусть объект $x \in D(q,r)$ при некотором разбиении попадал в обучение, и пессимистичный метод МЭР выбрал правило r. Тогда на этом же разбиении правило q будет иметь на одну ошибку меньше на обучении, так как правило r ошибается на всех объектах из $D(q,r)$, а q — нет. Значит метод МЭР должен выбрать q. Пришли к противоречию, следовательно, доказываемое утверждение верно. ■

Теорема 4.3 *Пусть множество объектов $D(q,r)$ состоит из одного объекта x. Тогда, чтобы пессимистичный метод МЭР выбрал правило q, объект должен находиться в X_q.*

Доказательство.

Докажем от противного. Пусть объект $x \in D(q,r)$ при некотором разбиении попадал в контроль, и пессимистичный метод МЭР выбрал правило q. Тогда на этом же разбиении правило r будет иметь на одну ошибку больше на контроле, так как правило r ошибается на всех объектах из $D(q,r)$, а q — нет. Значит метод МЭР должен выбрать r. Пришли к противоречию, следовательно, доказываемое утверждение верно. ■

Следствие. Объекты недоминирующегося подмножества S, такого, что $r(x,S) \sim r(x)$, являются фиксированными для правила r. Для правила r класса $y \in Y$ в X_r входят объекты из S класса y, а в X'_r — объекты из S всех остальных классов.

Рассмотрим снова граф связей правил (рис. 2), но теперь расположим правила снизу вверх, по возрастанию числа ошибок на полной выборке $m(r,\mathbb{X})$, см. рис. 3. В первом (нижнем) слое расположены правила с наименьшим числом ошибок (в данном примере есть корректное правило, не допускающее ошибок на полной выборке). В следующем слое расположены правила, допускающие на одну ошибку больше, и так далее.

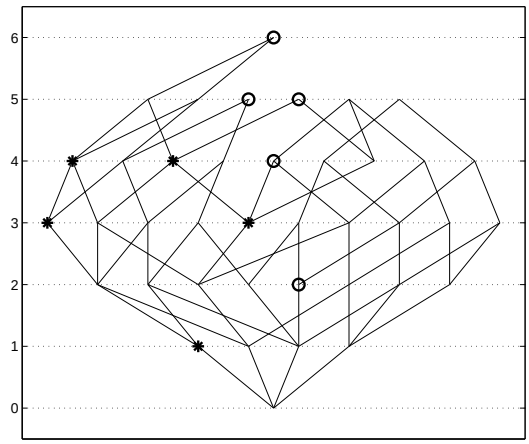

Рис. 3: Упорядоченный по величине ошибок граф связей. Граф изоморфен графу связей, изображенному на рис. 2. По вертикальной оси отложено число ошибок $m(r,\mathbb{X})$.

Теорема 4.4 *Пусть Q — множество правил, связанных с правилом r, таких, что $|D(q,r)| = 1$. Тогда*

$$\bigcup_{q \in Q} (X'_q \cup D(q,r)) \subseteq X'_r.$$

Доказательство.

Множество правил Q лежит слоем ниже относительно правила r. Утверждение теоремы можно переписать в виде:

$$\bigcup_{q \in Q} X'_q \;\cup\; \bigcup_{q \in Q} D(q,r) \subseteq X'_r.$$

Согласно теореме 4.2, $\bigcup\limits_{q \in Q} D(q,r) \subseteq X'_r$.

Если исключить из рассмотрения объект $x = D(q,r)$, то правила r и $q \in Q$ имеют одинаковые векторы ошибок, могут выбираться пессимистичным методом МЭР только на одинаковых разбиениях, следовательно, $X'_q \subset X'_r$ для всех $q \in Q$. ■

Для отыскания фиксированных точек воспользуемся следующим алгоритмом. Будем просматривать правила послойно в порядке увеличения числа ошибок на полной выборке.

Внутри слоя порядок просмотра правил не важен. Для каждого просмотренного правила r возьмем множество правил

$$Q = \{q \in R\colon |D(q,r)| = 1\},$$

то есть правила, связанные с r и находящиеся на слой ниже. Для каждого такого правила $q \in Q$ добавим $x = D(q,r)$ в X_q по теореме 4.3. По теореме 4.4 добавим в X'_r объекты множества $\bigcup\limits_{q\in Q}(X'_q \cup D(q,r))$.

Заметим, что в рамках теории Вапника-Червоненкиса возможно получить оценку с аналогичной структурой [5]:

$$Q_\varepsilon \leqslant \sum_{r\in R} H_L^{\ell,m_r}\big(\tfrac{\ell}{L}(m_r - \varepsilon k)\big),$$

где $m_r = m(r,\mathbb{X})$. Эта оценка учитывает расслоение семейства правил R по уровням числа ошибок m_r. Оценка (5) более точна за счет учета связности. Чем точнее будут верхние оценки вероятностей P_r, тем точнее будет и оценка (5).

5 Эксперименты и выводы

Для эксперимента на модельных данных положим: число признаков $n = 2$, число классов $Y = \{0,1\}$, число объектов $L = 100$, по 50 объектов в каждом классе. Возьмём четыре модельных выборки «Correct», «Noise10», «Noise20» и «Random», отличающихся только классификацией объектов. Для выборки «Correct» существует правило, разделяющее два класса без ошибок. Выборка «Noise10» получается из «Correct» небольшим зашумлением: для 10 пограничных объектов класс меняется на противоположный. Для выборки «Noise20» класс меняется у 20 объектов. Выборка «Random» получается случайным назначением классов всем объектам. В качестве ориентира используем оценки $\hat{Q}_\varepsilon$, вычисляемые методом Монте-Карло по 100 случайным разбиениям $\mathbb{X} = X \sqcup \bar{X}$, см. рис. 4.

Рис. 6 показывает, что завышенность оценки (5) тем больше, чем менее точной является закономерность, содержащаяся в выборке. Также это можно наблюдать на рис. 5.

Рис. 7 показывает, что завышенность полученных оценок относительно невелика.

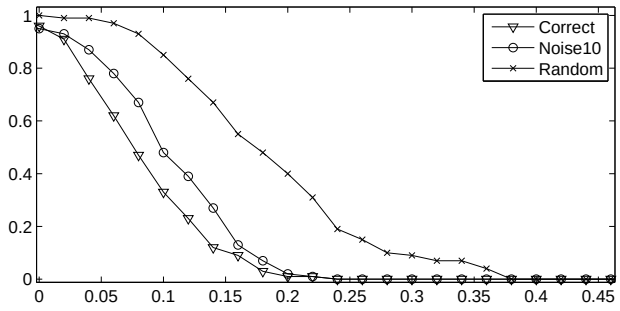

Рис. 4: Зависимость $\hat{Q}_\varepsilon$ от ε для трёх выборок.

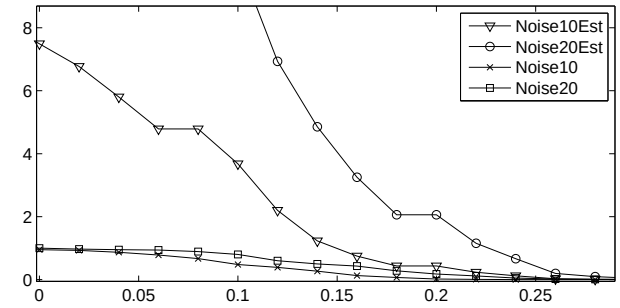

Рис. 5: Сравнение зависимостей $\hat{Q}_\varepsilon$ и Q_ε от ε для выборок «Noise10» и «Noise20».

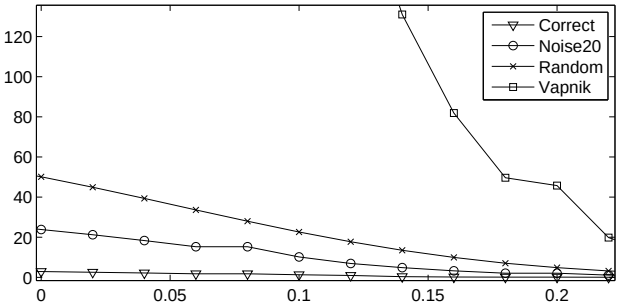

Рис. 6: Зависимость верхней оценки Q_ε, вычисленной по формуле (5), от ε, для трёх выборок.

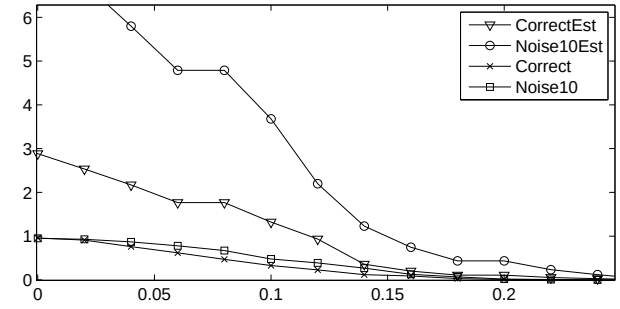

Рис. 7: Сравнение зависимостей $\hat{Q}_\varepsilon$ и Q_ε от ε для выборок «Correct» и «Noise».

Список литературы

[1] *Вапник В Н., Червоненкис А. Я.* Теория распознавания образов. — М.: Наука, 1974.

[2] *Донской В. И., Башта А. И.* Дискретные модели принятия решений при неполной информации. — Симферополь: Таврия, 1992. — 166 с.

[3] *Кочедыков Д. А., Ивахненко А. А., Воронцов К. В.* Применение логических алгоритмов классификации в задачах кредитного скоринга и управления риском кредитного портфеля банка // Всеросс. конф. ММРО-13, 2007. — С. 484–488.

[4] *Martin J. K.* An exact probability metric for decision tree splitting and stopping // Machine Learning. — 1997. — Vol. 28, No. 2–3. — Pp. 257–291.

[5] *Vorontsov K. V.* Combinatorial probability and the tightness of generalization bounds // Patt. Rec. and Image Anal. — 2008. — Vol. 18, No. 2. — Pp. 243–259.

[6] *Vorontsov K. V.* Splitting and similarity phenomena in the sets of classifiers and their effect on the probability of overfitting // Patt. Rec. and Image Analysis. — 2009. — Vol. 19, No. 3. — Pp. 412–420.

[7] *Vorontsov K. V.* Exact combinatorial bounds on the probability of overfitting for empirical risk minimization // Patt. Rec. and Image Analysis. — 2010. — Vol. 20, No. 3. — Pp. 269–285.